

SOCIAL MEDIA ANALYTICS

ABSTRACT

Since an ever-increasing part of the population makes use of social media in their day-to-day lives the sphere of social data so generated is increasing in all its dimensions with great velocity. Study about the strategies for the analysis of this huge amount of data, in order to harness the crucial information hidden in its greater depth, is done in various engineering and science disciplines. The social media analytics process involves four distinct steps viz. data discovery, collection, preparation, and analysis. A key analytical task is to study the social media sites for knowledge exploration and predictive analysis. Data is collected, parsed and analyzed for relationships, and features are selected and mapped to estimate the response of a system under exploration.

As social media data has grown abundantly, it can be captured and processed, to represent the potential behavioral patterns in society. In turn, this unstructured social media data can be parsed and integrated as a key factor for predictive analysis. This paper presents a comprehensive survey to characterize this fast-growing area and summarizes the state-of-the-art for analyzing social media data. In order to explore how predictions might be performed in such a framework, it also presents some results from a user study focusing on social media data as a predictor. The review focuses on different domains like industrial domains, data-mining objectives, use cases and applications. The review also offers insights into the analytical approaches that are being used in industrial domains for decision making.

KEYWORDS:

Data mining, topic modeling, predictive analysis, extracting, analyzing trends, behavioral patterns, social media, LDA, R programming, MongoDB.

INTRODUCTION:

Social media is a web and mobile based Internet applications that allow the creation, access and exchange of user-generated content that is easily accessible. Besides social networking sites (e.g., Twitter and Facebook), this paper uses the term ‘social media’ in a broader sense to encompass really simple syndication feeds, blogs, wikis and news, all typically yielding unstructured text and accessible through the web. Social media analytics is concerned with “developing and evaluating informatics tools and frameworks to collect, monitor, analyze, summarize, and visualize social media data. It facilitates the conversations and interactions in order to extract useful patterns and intelligence”

Research on social media has intensified in the past few years as it is seen as a means of garnering insight into human behaviors. The unstructured nature of social media data also provides unique challenges and opportunities for researchers across a variety of disciplines. Businesses are tapping into social media as a rich source of information for product design, relations management and marketing. Scientists utilize social media data as a platform for developing new algorithms for text mining and sentiment analysis and focus on social media as a sensor network for natural experimentation for exploring social interactions and their implications.

In this paper, we present a framework for social media integration, development, analysis and prediction. This framework consists of tools for extracting, analyzing and modeling trends across various social media platforms. In order to test our framework, we focus on the specific problem of predicting the user’s interest through our own social media site as well as other. This system integrates unstructured data from Twitter and Facebook with organized data from the insights. Temporal trends and sentiments are extracted and visualized from social media.

STATE OF THE ART:

Social network analysis has gained prominence due to its use in different applications-from product marketing to search engines and organizational dynamics. The basic motivation of this project is the demand to exploit knowledge from copious amount of data collected, pertaining to social behavior of users in online social sites. The following are some research papers and other literature works for gaining knowledge of big data, topic modeling, social media analysis, platforms and techniques to design the social media analytics system and data mining.

1. Text Mining with R: A tidy approach by, Julia Silge and Robert Davidson

This book serves as an introduction of text mining with a brief description about important libraries like the tidy text and the topic model packages in R. The functions provided by the tidy text package are helpful in transforming data into more intelligible form. It also provides compelling examples of real text mining problems. It explains how these methods implement the tidy data principles and make the text mining task easier, more effective, and consistent. Treating text as data frames of individual words allows to manipulate, summarize, and visualize the characteristics of text easily and integrate natural language processing into effective workflows already in use.

2. Social media analytics: a survey of techniques, tools and platforms by, Bogdan Batrinca, Philip Treleaven[1] :

It presents a comprehensive review of software tools for social networking media such as wikis, really simple syndication feeds, blogs and news groups. It also includes introductions to social media storage, data cleaning, sentiment analysis and methodology and a critique of social media tools. Analyzing social media, in particular Twitter feeds for sentiment analysis, has become a major research and business activity due to the availability of web-based application programming interfaces (APIs) provided by Twitter, Facebook and News services. This has led to an ‘explosion’ of data services, software tools for scraping and analysis and social media analytics platforms.

3. Social-Network Analysis Using Topic Models by Youngchul Cha, Junghoo Cho [2]:

This paper, explore techniques to provide more structure to a relationship by grouping the users based on their topic interests, and by labeling each relationship with the identified topic group. In particular it first studies the application of well-known Latent Dirichlet Allocation (LDA) model and its existing variants to the graph-labeling task and demonstrates that the existing models do not handle popular nodes (nodes with many incoming edges) in the graph very well.

4. Real-Time Detection of Traffic From Twitter Stream Analysis by Leonora D’Andrea, Pietro Ducange, Beatrice Lazzerini [5]:

This paper presents a real-time monitoring system for traffic event detection from Twitter stream analysis. The system fetches tweets from Twitter according to several search criteria; processes tweets, by applying text mining techniques; and finally performs the classification of tweets. The traffic detection system was employed for real-time, allowing for detection of traffic events almost in real time, often before online traffic news web sites.

METHODS:

Following are the methods that we have implemented in our project:

1. Design and development of social media website :

This project consists of two categories of social sites. The first one includes the existing social media websites such as Facebook and Twitter. In order to get more flexibility and to overcome the proprietary issues, a new social media website has been designed and developed. It helps us to gain full-fledged access to all the data as per the requirement for analysis.

2. Fetching and extracting data from social sites using API's :

The Application Program Interface (API) acts to integrate and enrich existing applications with social data. This module includes the use various API's specific to the social sites. The API's provide various features for fetching and manipulating the live social data. The API's that are used in this project are: Facebook (Graph API, RESTFB), Twitter (Twitter4j).

3. Storage of Raw data and Report:

This module consists of the database that will be required for the storage of the fetched data. This project contains both MySQL and MongoDB as the data stores for the fetched data. The database stores the raw data for analysis while the visual reports generated in analysis are stored as an image in to base file structure for simplicity.

4. Data Processing:

This is the module where the actual analysis of the data takes place. The raw data is converted into factual data to help categorizing the users according to the requirement. To extract the hidden pattern from the data the LDA algorithm is used. The result are stored as reports which are made available to the users by integrating them into our social media site itself. Latent Dirichlet Allocation (LDA) is a well-studied algorithm that discovers latent topics from a corpus of documents so that the documents can then be assigned automatically into appropriate topics. It studies the relationship among various sentiments and assigns them to each topic. More fascinating fact about LDA is its ability to divide each document into a given number of topics.

5. Topic modeling for users:

In this module customer insights are generated, categorizing them into various topics of their interest. It aggregates the analyzed information from page likes, shared posts and comments. It helps to categorize people according to their interests as well as to decide market strategies and deliver targeted messages to customers.

IMPLEMENTATION:

Algorithm:

LDA (LATENT DIRICHLET ALLOCATION)

Latent Dirichlet Allocation is one of the most common algorithms for topic modeling. Latent Dirichlet allocation (LDA) is a particularly popular method for fitting a topic model. It treats each document as a mixture of topics, and each topic as a mixture of words. This allows documents to “overlap” each other in terms of content, rather than being separated into discrete groups, in a way that mirrors typical use of natural language. There are a number of existing implementations of this algorithm, and two of them are depicted in the figure below.

Pseudocode of LDA based on Topic Modeling is as follows:

```
{  
    Document ← fetch (document from mongo)  
    C_document ← Corpus (document)  
    Tokens ← unnest_tokens(c_document)  
    dtm() ← Document Term Matrix(tokens)  
    lda_doc ← LDA(dtm,k)  
    plot ← (lda_doc)  
}
```


REFERENCES:

Research Papers:

1. B. Batrinca , P. C. Treleaven. (2014).” Social media analytics: A survey of techniques, tools and platforms,” AI & SOCIETY. 30. 89-116. 10.1007/s00146-014-0549-4.
2. Youngchul Cha, Junghoo Cho,” Social-Network Analysis Using Topic Models,” in SIGIR’12, 2012 .
3. Ashish Bindra, ”SocialLDA: Scalable Topic Modeling in Social Networks”, M.S Thesis ,Univ. of Washington,Seattle,USA,2012.
4. William Ribarsky, Xiaoyu Wang,Wenwen Dou,” Social Media Analytics for Competitive Advantage,” in EuroVis Workshop on Visual Analytics ,2013.
5. Eleonora D’Andrea, Pietro Ducange, Beatrice Lazzerini, ”Real-Time Detection of Traffic From Twitter Stream Analysis” in IEEE Transactions on Intelligent Transportation Systems, 2015.

Weblinks:

6. Alex Fox.(2013,April 26).Why Tracking your social media is important. [online].Available:<http://www.seguetech.com/why-tracking-your-social-media-analytics-is-important>
7. Steve Hamby.(2012,May 15).Five Reasons to Use Social Media Analysis. [online].Available:http://www.huffingtonpost.com/steve-hamby/social-media-analysis_b_1344666.html
8. P.Dasgupta, Rani Bansal. (2010). social Media Analytics.[online].Available : https://www.ciosummits.com/media/pdf/solution_spotlight/tcs_social-media-analytics.pdf
9. <https://www.tidytextmining.com/tidytext.html> (LDA algorithm)

CONCLUSION:

As organizations get more serious about measuring the effectiveness of their social media initiatives, it is important for them to link their measurement frameworks to high level business objectives of revenue generation, cost reduction, or operational excellence. Situating a social media analytics program as part of the overall business intelligence strategy provides a practical approach for organizations to get the most of their investments in such initiatives. Such an orientation would also enable organizations to improve the tactical execution of social media strategies such as building engagement across various channels and platforms, while directly tying social media analytics to their business objectives. The discussion in this paper can serve as a basis for establishing a baseline philosophy for businesses partaking various social media initiatives.